DOCUMENT RESUME

ED 446 997 SE 064 302

AUTHOR	Manning, Colleen F.; Goodman, Irene F.
TITLE	A Longitudinal Evaluation of the National Cancer Institute Science Enrichment Program.
PUB DATE	2000-00-00
NOTE	18p.; Produced by the Goodman Research Group, Inc.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; Enrichment Activities; *Evaluation Methods; *Longitudinal Studies; Program Evaluation; Qualitative Research; *Science Education; Statistical Analysis

ABSTRACT

This paper proposes the design and key methodological features of a longitudinal evaluation of the National Cancer Institute Science Enrichment Program (NCISEP). Goodman Research Group's (GRG) five-year longitudinal evaluation is designed as a randomized experiment with a control group and employs both quantitative and qualitative data collection methods. It states that given that SEP's goal of influencing career development is a long-term one, it is necessary that the evaluation of such a project be designed as a longitudinal study. Moreover, in order to attribute effects to the intervention, the study must include a control group of students who do not attend the program. Finally, it concludes that quantitative and qualitative data collection methods are vital to ensuring an in-depth understanding of how a program achieves its goals. (ASK)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 446 997

A Longitudinal Evaluation of the National Cancer Institute Science Enrichment Program

Colleen F. Manning, Senior Research Associate & Irene F. Goodman, President

Goodman Research Group, Inc.

INTRODUCTION

This paper focuses on the design and key methodological features of a longitudinal evaluation of the National Cancer Institute Science Enrichment Program (NCI SEP). Goodman Research Group's (GRG) five-year longitudinal evaluation is designed as a randomized experiment with a control group, and employs both quantitative and qualitative data collection methods. Five cohorts of SEP students (i.e., students attending SEP in summers 1998-2002) and two cohorts of control group students (i.e., students recruited into the control group in summers 1999 and 2000) will comprise the evaluation sample.

The Office of Special Populations Research (OSPR), within the office of the NCI Director, administers SEP. OSPR developed the intervention program to respond to the problem of underrepresentation of biomedical scientists from minority and underserved populations. The program serves rising tenth grade high school students from minority and underserved populations with the primary goal of encouraging their interest in a science, mathematics, or research career. NCI also seeks to broaden and enrich students' sociocultural backgrounds. SEP is a five to six-week summer residential program currently taking place on two university campuses. Each regional program serves about 50 students per summer.

SEP has a 10-year history. In 1990 and 1991, SEP pilot programs took place at Hood College in Maryland. In 1992, NCI awarded contracts for SEP programs to four regional sites, where the programs ran through 1997. In 1998, NCI began a new SEP contract cycle and awarded 5-year contracts to the two current regional programs. Each of these two sites administered programs in

SE 064 207

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

C. Manning

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

1998 and 1999. With the new contract cycle, NCI also contracted with Goodman Research Group, Inc. (GRG) to serve as the SEP program evaluators for the five-year program cycle.

NCI SEP Evaluation Methodology

In order to make sound arguments about SEP's effectiveness, the program's evaluation includes three key methodological features: a longitudinal design, a randomized control group, and the use of quantitative and qualitative data collection components. Each of the methodological features is discussed in this paper.

Longitudinal Design

SEP's major goal of encouraging students to select a career in science, mathematics, or research is long-term in nature. Therefore, it is necessary to follow the program participants over time to determine the effectiveness of the program in meeting that goal. In addition to collecting data from students during the summer SEP, we collect data from students twice each year.

At the end of the five-year evaluation period, the first group of SEP students (i.e., Summer 1998) ought to be sophomores in college. Sophomore year in college is the furthest point along their educational career SEP students can be tracked, and the evaluators will be able to track only the first group of students this far.

This time line has implications for assessing SEP's effectiveness in meeting their goal. Obviously, this evaluation will not follow students until the point of career selection. However, the evaluation can and will assess *precursors* to selecting a career in one of these fields. We have defined these precursors as *interest in* and *preparation for* a career in one of these fields.

During their first year in the study, each cohort will complete pre- and post-tests designed to assess the major areas of interest to the evaluation: attitudes about science and math, career

aspirations and expectations, and science process skills. Students will be tracked and surveyed on an annual basis thereafter. It will be possible to follow the first two SEP cohorts and one of the two control group cohorts into college. This opportunity is perhaps the most critical aspect of the evaluation plan because it makes possible the investigation of SEP's longer-term goal.

We have completed summer pre- and post-testing with the SEP 1998 and 1999 students and with the control group 1999 students, bringing our total SEP sample to 183 students. We also have conducted the first annual follow-up survey of the SEP 1998 cohort. In June 2000 we will conduct the second annual follow-up of this cohort, as well as the first annual follow-up of the SEP and control group 1999 cohorts. Table 1, on the following page, provides an overview of the longitudinal design and our progress to date.

Table 1

Overview of NCI SEP Evaluation*

Data Collection	Pre-tests	Post-tests	1st Follow-up Postcard	1st Follow-up Survey	2nd Follow-up Postcard	2nd Follow-up Survey	3rd Follow-up Postcard	3rd Follow-up Survey	4th Follow-up Postcard	4th Follow-up Survey	5th Follow-up Postcard
Students' grade level at point of data collection	*Completed 9th grade*	*Rising 10th grader*	*Mid 10th grade*	*Post 10th grade*	*Mid 11th grade*	*Post 11th grade*	*Mid 12th grade*	*Post 12th grade*	*Mid Frosh year*	*Post Frosh year*	*Mid Sophomore year*
Cohort 1: 1998 SEP students	June 1998	August 1998	January 1999	June 1999	January 2000	June 2000	January 2001	June 2001	January 2002	June 2002	January 2003
Cohort 2: 1999 SEP & Control students	June 1999	August 1999	January 2000	June 2000	January 2001	June 2001	January 2002	June 2002	January 2003		
Cohort 3: 2000 SEP & Control students	June 2000	August 2000	January 2001	June 2001	January 2002	June 2002	January 2003				
Cohort 4: 2001 SEP students	June 2001	August 2001	January 2002	June 2002	January 2003						
Cohort 5: 2002 SEP students	June 2002	August 2002	January 2003								

*Shaded areas indicate completed data collection points.

Two especially important considerations for longitudinal studies are response rates and long-term data management. Our response rate to the summer pre- and post-tests has been 100% for the treatment group. The first annual follow-up survey with the 1998 cohort in June 1999 yielded a response rate of 74%. It is our policy to follow up with non-respondents both by re-sending surveys and by sending several reminder postcards. We find these to be effective strategies, increasing our response rates by 10% on average.

Decisions about data management are paramount to longitudinal studies. GRG has developed a SEP evaluation database in Microsoft Access, a relational database management system for Microsoft Windows. Information about the two programs is stored in one table, while information about the students is kept in a separate table. However, all the tables are related to one another so that data from the different tables may be combined for data analysis and reporting purposes.

The purpose of the database is to store information for mailing, tracking, and data analysis purposes. Each SEP student and control group student has been assigned an identification (ID) number. The ID number appears in the database and on every survey that the student receives. This allows us to easily track non-respondents and send them reminders to return their surveys. Each year, information on every student will be entered into the database. This information includes the student's name, address, phone number, e-mail, birth date, program, and SEP year.

Randomized Control Group

Perhaps the biggest challenge we faced in designing the NCI SEP evaluation was incorporating a suitable comparison or control group. Although involving a control group of students who do not attend SEP in a longitudinal study is time-consuming and costly, it will be the only way to know the extent to which observed SEP effects can be attributed to the program. Often, programs will compare their own statistics to national (or local) trends as a means of assessing their success. While such statistical comparisons are interesting and can be informative, it is not possible to

attribute any differences between national trends and program trends to programmatic interventions.

A comparison of SEP students to the general population would not be scientifically sound given the way that SEP students are selected for participation in the program. The students are self-selected by their interest in and motivation to attend the program. Many are encouraged by their science teachers to apply to the program because of their demonstrated interest and/or ability in science. Clearly, these students are different from the general population on the outcomes of interest to the evaluation.

Another possibility for a control group that we rejected was using students enrolled in other science enrichment programs. These students are similar to the SEP students in terms of their interest in science and their initiative to enroll in a science program. However, this option is undesirable for a couple of reasons. First, our research into other science enrichment programs indicated they were different from SEP in a variety of ways (e.g., different target audience, focus, length, or format), and controlling for those differences would be virtually impossible. Second, it would be very difficult to control for differences between SEP students and students from other programs in terms of background variables, such as geography and type of high school.

Ultimately, we proposed for NCI's consideration two controlled study designs: a randomized experimental design and a quasi-experimental design (Cook and Campbell, 1979). Each of the design options had its advantages and disadvantages, as summarized in Table 2.

Table 2

Summary of Proposed Design Options and Samples

	Treatment Group Selection	Comparison Group Selection	Pros	Cons
Option 1: Randomized Experimental Design	One-third selected by program + two-thirds randomly assigned by evaluator (from a pool selected by program)	All randomly assigned by evaluator	• Provides the most firm conclusions about SEP's effectiveness	• Requires program to use a modified selection procedure • Requires large applicant pool • Smaller control group than Option 2
Option 2: Quasi-Experimental Design	All selected by program	All matched teachers and selected by evaluator	• May be more practical for programs • Larger control group than Option 1	• Less definitive attributions of causality than Option 1

The NCI Program Officer, the two SEP Program Directors, and GRG agreed on the desirability of the stronger of the two research designs, the randomized controlled experiment. The procedure for the randomization is as follows:

Step 1: In 1999 and 2000, each SEP program selects up to one-third of their student body from their applicant pool. This step ensures that the programs have an opportunity to accept those students in whom they are most interested as well as those representing geographic areas and high schools of particular concern to the program.

Step 2: From among the remaining applicants, the programs over-select students for admission. That is, the SEPs select at least twice as many students as they need to fill the remaining two-thirds of their student body.

Step 3: The evaluators randomly assign these students to the treatment group (i.e., they are offered admission to the program) or to the control group (i.e., they are not admitted, and the evaluators recruit them to participate in the control group).

Step 4: The programs inform the control group students that they have not been accepted into the program. The evaluators separately contact and recruit the potential control group students, informing them that they have been selected to participate in a research study of students who are interested in science and science programs. The students are informed that the SEP program referred us to them.

We recruited the first cohort of control group students in Spring/Summer 1999. After following the procedure outlined above, the potential pool of control group students from both regional programs was 87. Of the 87, 75 (86%) agreed to participate in the control group; 49 of the 75 (65%) completed both pre- and post-test surveys.

To begin with, the potential pool of 87 was smaller than GRG had expected; one of the two regional programs in particular had underestimated the challenge of the enhanced recruiting. If we recruit another 50 or so students into the control group this Spring/Summer, the total control group sample will be approximately 100 students. While GRG had planned only to recruit two cohorts of control group students (in 1999 and 2000, years two and three of the study), because of the length of the study and expected attrition, GRG and NCI are discussing the possibility of recruiting an additional cohort of control group students (in 2001, year four of the study).

Preliminary analysis indicates that the treatment group (i.e., only those SEP students who were randomly assigned to the programs by the evaluator) and control groups are equivalent in all regards studied except gender. Because of a dearth of male applicants to one of the two regional programs, the control group contains more females than males. Table 3 offers a brief profile of the treatment and control groups to date. The data show the groups to be similar in terms of race, first language, previous science and academic experience, and parental support for science/math.

Table 3
Student Profile: SEP versus Control Group

		SEP	Control
Gender ($p<.05$)	Female	52%	70%
	Male	48	29
Race	American Indian	8	3
	Asian	2	3
	Black	22	23
	Native Hawaiian or Pacific Islander	3	8
	White	33	27
	Latino/Hispanic	20	21
	Other	13	15
English first language	Student	81	70
	Mother	67	64
	Father	68	62
Previous science & other academic activities	science program not on college campus	14	10
	science program on college campus	17	10
	college course	9	14
	science fair	73	68
	research	14	7
	health care	9	18
	after-school academic club	44	48
	tutored	53	51
Support for science & math at home	talk about what they're learning in science or math	76	74
	help with homework	65	66
	help with project	68	59
	show how to do experiment	32	23
	show how to do problem	60	72
	watch science on TV	35	32
	talk about science or math topics	63	57

Quantitative and Qualitative Components

The study is grounded in the premise that advancing knowledge about SEP requires the capability to generalize about the two sites while remaining sensitive to their individual contexts. Therefore, we include quantitative and qualitative data collection components, both equally important in ensuring such capability. For the purposes of this paper we limit our discussion to the quantitative student measures. The qualitative portion of the study includes annual site visits to each of the two programs.

The evaluation measures have been developed and/or chosen in consultation with NCI staff, and each of the Program Directors had an opportunity to review the instruments. Gathering information from the students is accomplished via written surveys, which we find to be the most cost-effective means of collecting extensive quantitative data.

The pre- and post-test surveys are administered to SEP students on the first and last days of the program, respectively (with the exception of the pre-test attitude survey, which is mailed to students three to four weeks prior to SEP). The follow-up survey, accompanied by a letter and a postage-paid business reply envelope, is mailed by the evaluator to students. All surveys are mailed to the control group students at the same points in time they are mailed or administered to SEP students.

Student Attitude Surveys

There are two similar versions of each of the attitude surveys: one for SEP students and a slightly modified version for control group students. In developing the surveys we reviewed all of the surveys used by the 1991-1997 SEP programs, incorporating some of their questions into our instruments. We also based some of the questions on previously developed and tested measures (e.g., Fennema and Sherman, 1976).

The surveys reflect the goals of SEP and are designed to assess changes in students' attitudes

about and experiences with science, mathematics, and research. The pre-test survey serves as a baseline measure. The post-test survey allows us to assess changes in students over the five- to six-week SEP period, and the follow-up surveys will investigate whether changes are sustained over time.

The surveys obtain the following categories of information:

1) background information (*pre-test only*) — including, but not limited to, demographic information (e.g., age, gender, race/ethnicity), information about their home life (e.g., parents' educational level and occupation), information about their 9th grade year (e.g., what courses they took, what the instruction was like, how much homework they were assigned, the grades they received, how much they enjoyed their classes, how challenging they found the classes), and extracurricular information (e.g., participation in programs/clubs/activities, jobs, volunteer work).

2) science and math information (*pre-test, post-test, and follow-up*) — including students' motivation with regard to science and math and their interest in and attitudes about science and math;

3) career information (*pre-test, post-test, and follow-up*) — including their career aspirations and expectations, their knowledge about the necessary preparation (e.g., years and type of education) for their career, and their awareness of different types of science, mathematics, and research careers; and

4) SEP information (*pre-test, post-test, and follow-up*) — The pre-test for SEP students includes questions about how they heard about SEP, why they applied, and their expectations of the program. (Control group students are asked whether they heard of or knew anyone who attended SEP.) The post-test for SEP students contains rating scales for various aspects of the program, including the academic curriculum (e.g., student ratings of the effectiveness of instruction), resources/materials (e.g., laboratory equipment,

computer applications), seminars, lectures, field trips, and cultural events. It also includes a couple of open-ended questions.

Although a core set of questions about students' attitudes remains the same from the pre-test survey through all the follow-up surveys, the follow-up surveys are and will be tailored to respond to the grade level of the cohort to whom they are being sent. For example:

- The first and second follow-up surveys sent to students in the 10th and 11th grades, respectively, will obtain information about whether students are electing to take science and math courses, and their participation in extracurricular science and math activities.

- The third follow-up survey sent to students in the 12th grade will ask about their SAT scores and their college plans.

- The fourth and fifth follow-up surveys sent to students in their frosh and sophomore years of college, respectively, will include questions about science and math course taking, and intended or declared major.

A follow-up postcard is sent to each cohort annually in January. Although it does contain a few brief questions, its primary purpose is to assist the evaluator with tracking. Our experience has shown us that the more contact the evaluator has with the sample, the less attrition there is. This has and will also enable us to better track changes in student information, such as change of addresses.

Test of Integrated Process Skills

The SEP program does not have a specific science curriculum per se; instead, it aims to develop in students skills that are universal to all disciplines of science. Therefore, we are using previously developed and validated tests to assess changes in students' science *process* skills. We are using a combination of *The Test of Integrated Process Skills* (TIPS and TIPS II) (Dillashaw and Okey,

1980; Burns, Okey, and Wise, 1985) and the *Test of Logical Thinking* (TOLT) (Tobin and Capie, 1981).

The TIPS is a 36-item science process skill test for middle and high school students. It takes approximately 45 minutes to complete. There are three key advantages to using the TIPS in the SEP evaluation:

1) Its content validity and reliability have been established. The TIPS is one of only a few process skills tests for middle and high school students that has gone through a rigorous test development process, with attention to content validity, reliability, difficulty and discrimination indices, response format, reading level, and item context. In addition, the test was reviewed by six experienced science educators.

2) It is not curriculum-specific. Given that the two regional SEPs have different science curricula, it is essential that any science measures used with both programs be non-curriculum-specific. The five process skill objectives covered by the TIPS include identifying variables, operationally defining, stating hypotheses, graphing and interpreting data, and designing investigations.

3) There are two versions of the test. The TIPS and the TIPS II are related to the same objectives, produce highly similar mean scores, have the same average difficulty index, and scores on the two tests are highly correlated. Together, they offer a total of 72 items for process skills assessment, making it very feasible to use two 36-item equivalent tests for pre- and post-assessment.

The TOLT was designed to measure five modes of formal reasoning: controlling variables, proportional reasoning, combinatorial reasoning, probabilistic reasoning, and correlational reasoning. It has been used with middle and high school students and has a high test reliability. It is a 10-item test with both multiple choice and open-ended questions that takes approximately 10 minutes to complete.

Data Analysis and Preliminary Results

The outcomes of interest in the evaluation are events in the lives of individual students. The following outcome variables will be considered in our final analysis:

- ❒ student interest in, and preparation for a major or career in, science and/or math, as demonstrated by course choices, involvement in extracurricular science/math activities, etc.;
- ❒ student belief that they will major in science or math; and
- ❒ student belief that they will have a career in science or math.

A number of variables having to do with individual students will be used in the analysis as predictors of the above outcomes or as control variables to adjust for differences between individual students. These include their parents' education, ethnicity, science/math competency, student satisfaction with their high school experience, and student self confidence in science.

The importance of choosing the right analytic techniques and having adequate statistical power is crucial in ensuring that we reach the right conclusions based on our data. Our approach to analyzing the rich data to be gathered in this investigation will be to use the data analytic strategy of Hierarchical Linear Modeling (HLM) (Bryk, Raudenbush,& Congdon, 1996). In this model, analyses will examine growth overall and growth within students. This model also allows for unevenly spaced data collection points. For all quantitative data, we will also run frequencies and significance tests, such as standard chi-squares, paired T-tests and correlations.

In Conclusion

Given that SEP's goal of influencing career development is a long-term one, it is necessary that the evaluation of such a project be designed as a longitudinal study. Moreover, in order to attribute effects to the intervention, the study must include a control group of students who do not attend the program. Finally, quantitative and qualitative data collection methods are vital to

ensuring an in-depth understanding of how a program achieves its goals. The final results of the evaluation will offer conclusive evidence of the effectiveness of this science enrichment program. We believe the sound design presented in this paper may benefit other researchers seeking to evaluate programs with long-term goals.

References

Bryk, A.S., Raudenbush, S.W., & Congdon, R. (1996). HLM 2/3. Chicago: Scientific Software.

Burns, J. C., Okey, J. R., & Wise, K. C. (1985). Development of an integrated process skill test: TIPS II. *Journal of Research in Science Teaching*, **22**, 169-177.

Cook, T. D. and Campbell D. T. (1979). Quasi-Experimentation: Design & Analysis Issues for Field Settings. pp. 103-117. Boston: Houghton Mifflin Company.

Dillashaw, F. G., & Okey, J. R. (1980). Test of the integrated science process skills for secondary science students. *Science Education*, **64**, 601-608.

Fennema, E. & Sherman, J. A. (1976). Fennema-Sherman Mathematics Attitude Scales.

National Science Foundation. Women, minorities, and persons with disabilities in science and engineering: 1998. Arlington, VA, 1999.

The Regents of University of California and The Trustees of the California State University. (1986). Algebra Readiness Test.

Tobin, K. G., & Capie, W. (1981). The development and validation of a group test of logical thinking. *Educational and Psychological Measurement*, **41**, 413-423.

U.S. Department of Commerce, Bureau of the Census (1992). *Education in the United States, 1992 Census of the Population*: CP-3-4.